Final Abstract Number: 43.007
*Session: Infectious Disease Surveillance I*
*Date: Thursday, April 3, 2014*
*Time: 12:45-14:15*
*Room: Ballroom*

## Design of a text-processing algorithm for extracting data on signs and symptoms from free text electronic medical records

M. Chen

*National University of Singapore, Singapore, NA, Singapore*

**Background:** Electronic medical records (EMR) are a potential source of syndromic data for surveillance, but require text processing algorithms to extract relevant data.

**Methods & Materials:** We analysed free text EMR for a predefined set of 28 signs and symptoms (Box). We used a "compiler approach" to transform EMR text by first tokenizing source text, and then parsing these tokens according to some grammar rules. Patterns and grammar rules were created from a pilot set of medical records from 50 episodes of care. Patterns for each sign/symptom include the sign/symptom itself, common misspellings, and synonyms. Patterns for keywords were also created to capture contextual features. To define the presence of a sign or symptom, we want to know if a symptom is either affirmed, negated or not mentioned. To do so, we had to capture keywords that will change the status of a symptom (e.g. negative words such as "no", "nil", "denies") and conjunction keywords such as "and", "or", ",", "/" which are used to chain a list of symptoms together.

Manual coding by a nurse familiar with EMR shorthand and abbreviations was used as the gold standard against which text processing algorithm was compared.

| | |
|---|---|
| Abdominal distension | Lymphadenopathy |
| Abdominal pain | Myalgia |
| Anorexia | Nausea |
| Chest pain | Night sweats |
| Chills | Orthopnea |
| Cough | Rebound |
| Diarrhea | Rhinorrhea |
| Dyspnea | Sneezing |
| Edema | Sore throat |
| Fatigue | Sputum |
| Fever | Tachypnea |
| Headache | Vomiting |
| Hemoptysis | Weight loss |
| Jaundice | Wheezing |

Box: List of signs and symptoms considered

**Results:** The results for a section of the raw EMR text and the processed result, highlight how the algorithm could accurately classify a symptom appearing in the text as either affirmed (e.g. "myalgia" and "fever", which were present "x1/7", i.e. 1 day), or negated (e.g. no chest pain) and deal with misspelt words and abbreviations. For instance, the algorithm recognises the string "-no ST/RN/cough" to refer to all three symptoms of sore throat, rhinorrhea and cough being negated.

In the 50 episodes of care analysed, there were 107 signs and symptoms "negated" and 179 "affirmed" by manual coding (Table 1). Machine classification achieved a sensitivity of 97% for detecting signs/symptoms that are "affirmed" on manual coding; specificity was 96% "negated" signs/symptoms as the negatives, and > 99% when including as negatives both signs/symptoms that are "negated" and absent by manual coding.

**Conclusion:** Our algorithm is potentially useful for syndromic surveillance.

| Table 2A: Counting signs/symptoms that are "negated" as negative | | | |
|---|---|---|---|
| | | Human coding | |
| | | Negated | Affirmed |
| Machine coding | Negated | 107 | 8 |
| | Affirmed | 3 | 179 |
| **Table 2B: Counting signs/symptoms that are "negated" or absent as negative** | | | |
| | | Human coding | |
| | | Negated / Absent | Affirmed |
| Machine coding | Negated / Absent | 1210 | 8 |
| | Affirmed | 3 | 179 |

http://dx.doi.org/10.1016/j.ijid.2014.03.714

**Type: Poster Presentation**

Final Abstract Number: 43.008
*Session: Infectious Disease Surveillance I*
*Date: Thursday, April 3, 2014*
*Time: 12:45-14:15*
*Room: Ballroom*

## Development and evaluation of a multiple-locus variable-number tandem-repeats analysis assay for subtyping *Salmonella typhi* strains from Sub-Saharan Africa

N. Tau, A.M. Smith, K.H. Keddy*

*Centre for Enteric Diseases, National Institute for Communicable Diseases; Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa*

**Background:** Typhoid fever a major public health problem in developing countries. Epidemiological investigations of *Salmonella enterica* serovar Typhi (*Salmonella* Typhi) infections using molecular sub-typing methods are challenging due to the highly homologous nature of *Salmonella* Typhi strains. In recent years, several approaches in using multiple-locus variable-number tandem-repeats analysis (MLVA) for molecular sub-typing of *Salmonella* Typhi have been made. To date, a standardized set of variable-number tandem-repeats (VNTR) loci for the typing of homologous *Salmonella* Typhi strains has not been established.

The aim of our study was to develop and evaluate a MLVA assay consisting of 5 VNTR markers to analyse *Salmonella* Typhi strains from Sub-Saharan Africa (SSA).

**Methods & Materials:** To develop and evaluate the MLVA assay, 50 *Salmonella* Typhi strains from humans were selected from the culture collection at the Centre for Enteric Diseases (CED) at the National Institute for Communicable Diseases (NICD) from a potential 1080 isolates. This strain panel represented the diverse *Salmonella* Typhi pulsed-field gel electrophoresis (PFGE) profiles in the CED database, specimen collection dates and geographic areas within the SSA region. The 50 *Salmonella* Typhi strains were used to evaluate 12 polymorphic VNTR loci that were previously published. These include TR1, TR2, TR3, TR4, TR5, Sal02, Sal06, Sal10, Sal16, Sal20, TR4500 and TR4600. The assay included PCR amplification of VNTR loci using fluorescently labelled primers and size determination of PCR products by capillary electrophoresis.

**Results:** Of the 12 VNTR loci that were evaluated, 5 (TR3, TR4, TR5, Sal06, Sal10 and TR4500) were found unsuitable as they showed poor allele variation (between 1 and 3 alleles). The remaining 6 VNTR loci showed good allele variation and good diversity indices; with Sal20, Sal 16 and TR1 having 6, 9 and 11 alleles

respectively; and Sal02, TR4699 and TR2 having 15, 20 and 23 alleles respectively.

**Conclusion:** In summary, we have identified 6 polymorphic VNTR loci suitable for MLVA analysis of *Salmonella* Typhi strains. The five most diverse VNTR loci will be selected for the MLVA assay and will be used to analyse *Salmonella* Typhi strains from SSA. This work will assist in rapidly identifying strain relatedness and assist outbreak detection in typhoid fever outbreaks.

http://dx.doi.org/10.1016/j.ijid.2014.03.715

**Type: Poster Presentation**

Final Abstract Number: 43.009
*Session: Infectious Disease Surveillance I*
*Date: Thursday, April 3, 2014*
*Time: 12:45-14:15*
*Room: Ballroom*

## Co-infection of malaria and influenza viruses in Uganda: A pilot study

B. Namagambo[1,*], J.T. Kayiwa[1], T. Byaruhanga[1], N. Owor[1], I. Nabukenya[1], B. Bakamutumaho[2], J.J. Lutwama[3]

[1] *Uganda Virus Research Institute, Entebbe, Uganda*
[2] *Uganda Virus Research Institute, Kampala, Uganda*
[3] *Uganda Virus Research Institute, Entebbe, Uganda*

**Background:** Influenza is a highly contagious viral infection of the respiratory passages causing fever, headache, severe aching, cough, and periodically causes epidemics especially in temperate environments. In many African countries, people do not visit clinics or hospitals with just influenza infection. Majority of ILI patients do not seek medical care and very few of those who do, get tested for influenza. The many of influenza cases that are reported by surveillance programs are from sentinel sites where patients come in with other medical problems. Clinically, influenza is not distinguishable from most other infectious diseases with fever in the tropics. Malaria is an important infectious disease and is still thought to be the main cause of febrile episodes. Most fevers are thought to be malaria. Our investigations sort to establish information on incidence of malaria in patients who are positive with Influenza infection.

**Methods & Materials:** This cross-sectional pilot study examined incidence of malaria among outpatient visits and hospitalizations associated with Influenza like Illnesses (ILI) and Severe Acute Respiratory Illness (SARI) during the period February 2011-November 2013 in children, youth and adults attending six health facilities of; Kawaala health centre III, Kitebi Health centre III, UVRI Clinic, Entebbe Hospital and Mbarara Regional referral hospital in Uganda. Nasopharyngeal and oralpharyngeal swabs were collected from patients meeting the WHO case definition for ILI and SARI. Influenza viruses were screened for using RT- PCR and the clinical data presenting diagnosis of malaria was collected and analyzed.

**Results:** Out of the 1020influenza specimens collected from cases; 754 (73.9%) patients were diagnosed with malaria; 116 (15%) of 754 were positive for Influenza and 638 (84.6%) were negative; positive for Influenza A were 71(9.4%) with two subtypes; 56(7.4%) A(H3) and 15(2.6%) Pandemic A(H1N1) 2009, and 45 (6.0%) were Influenza B viruses. Of the 116 positives 108(93.1%) were ILI and 8(6.9%) were SARI patients. Although 107 (92.2%) Children diagnosed with Malaria had Influenza, 9 (7.8%) Youth had Influenza whereas there was no Influenza in Adult.

**Conclusion:** Our data shows a high incidence of Influenza in children diagnosed with malaria.

http://dx.doi.org/10.1016/j.ijid.2014.03.716

**Type: Poster Presentation**

Final Abstract Number: 43.010
*Session: Infectious Disease Surveillance I*
*Date: Thursday, April 3, 2014*
*Time: 12:45-14:15*
*Room: Ballroom*

## *Clostridium difficile* infection in University Hospital Trnava: A hospital-based survey

A. Streharova[1,*], M. Garabasova[2], M. Betkova[3], L. Macekova[4], J. Brnova[5]

[1] *University Hospital Trnava, Trnava, Slovakia*
[2] *Trnava University, School of Health Sciences and Social Work, Trnava, Slovakia*
[3] *Faculty of Social Sciences and Health Care,Constantine the Philosopher University, Nitra, Slovakia*
[4] *Analytx s.r.o., Trnava, Slovakia*
[5] *Trnava University, School of Public Health Social Sciences, Trnava, Slovakia*

**Background:** Clostridium difficile infection (CDI) have increased in frequency and severity over the past decade and are a leading cause of hospital acquired infections. In Slovakia, mandatory surveillance of Clostridium difficile infection and others health care associated infection was established due to electronically national Epidemiological information systems (EPIS). But only a few cases of CDI were reported as a healthcare associated infection in EPIS during last years. Therefore the aim of this hospital-based survey was to assess incidence, risk factors and outcome of CDI in University Hospital Trnava, Slovakia in period 2010-2012.

**Methods & Materials:** We analysed all patients with laboratory confirmed CDI in University Hospital Trnava during three years period (RIDA®QUICK *Clostridium difficile* Toxin A/B immunochromatographic rapid assay for the qualitative determination of the toxins A and B of *Clostridium difficile* in stool samples, R-Biopharm). Incidence rate per 10 000 hospitalized patients were calculated and patients characteristic were recorded from hospital information systems.

**Results:** Together 208 hospitalized patients were confirmed *Clostridium difficile* toxin in stool samples. Incidence of CDI in hospitalized patients during three years period 2010, 2011, 2012 increased 24/10 000, 27/10 000, 32/10 000 respectively. Health care –associated CDI were more often identified than community-acquired CDI (73,6% vs. 26,4%). Mean age of infected patients were $73 \pm 16$ (range 15-96) and female (60%) were more frequent infected than male. The most affected were Geriatric Department (29,3%), Department of Infectious Diseases (25%) and Department of Internal Medicine (24%). Concerning risk factors, most patients before development CDI received antibiotics - ciprofloxacin, cefuroxime and ampicillin/sulbactam due to respiration or urinary tract infection. Recurrence of CDI was confirmed in 7,7% patients and occurring within 3 months after the first episode. Almost all cases (99%) were treated with metronidazole, others 4 cases with rifaximin (2), vancomycin (1) and fidaxomicínom (1). Death associated with CDI was observed in two patients.